DATA VISUALIZATION WITH COVID DATASET

A PROJECT REPORT

Submitted by,

AIYAAN HASAN - 20201ISE0092
ABU SUFIYAN- 20201ISE9002
TUSHAR TIWARI - 20201ISE0091

Under the guidance of,
Ms. POORNIMA S

in partial fulfillment for the award of the degree of
BACHELOR OF TECHNOLOGY

IN

INFORMATION SCIENCE AND ENGINEERING

At

PRESIDENCY UNIVERSITY
BENGALURU
JANUARY 2024

ABSTRACT

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has emerged as one of the most significant global crises in recent history, profoundly impacting public health, economies, and societies worldwide. As the pandemic continues to unfold, there is an urgent need for comprehensive analysis and visualization of its spread and effects to inform decision-making and response efforts.

In this project, we present a thorough examination of the COVID-19 pandemic using extensive data sourced from the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE). Our analysis employs sophisticated data visualization techniques to provide deep insights into various aspects of the pandemic, spanning epidemiological trends, geographical distribution, and temporal dynamics.

At the outset, we offer a comprehensive snapshot of the current state of the pandemic on a global scale, elucidating the total number of confirmed cases, deaths, and recoveries. By calculating key epidemiological metrics such as mortality rate and infection rate, we provide a nuanced understanding of the severity and transmissibility of the virus across different regions.

To gain insights into the spatial spread of COVID-19, we leverage geospatial visualizations to map the distribution of cases and fatalities across countries. By visually depicting hotspots and patterns of transmission, these visualizations highlight disparities in the impact of the pandemic and underscore the importance of targeted interventions and resource allocation.'

Moreover, we conduct in-depth analyses of the top 10 worst-hit countries, examining the factors contributing to their disproportionate burden of cases and deaths. Through dynamic visualizations such as bar chart races, we track the evolution of COVID-19 cases and deaths over time, providing a compelling narrative of the pandemic's trajectory and the effectiveness of containment measures.

Additionally, we explore the temporal dynamics of COVID-19 through time-series analysis, identifying trends and patterns in case counts, mortality rates, and recovery rates. By visualizing these trends over time, we offer valuable insights into the effectiveness of public health interventions and the emergence of new challenges and opportunities in the fight against the pandemic.

Overall, our project contributes to a deeper understanding of the COVID-19 pandemic by leveraging data visualization techniques to synthesize complex information into accessible and actionable insights. By shedding light on the multifaceted dimensions of the pandemic, we aim to support informed decision-making, resource allocation, and policy development efforts to mitigate its impact and safeguard public health and well-being.

TABLE OF CONTENTS

CHAPTER-1
INTRODUCTION

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has sparked a global crisis of unprecedented proportions, profoundly affecting public health, economies, and societies worldwide. Since its emergence in late 2019, the virus has rapidly spread across continents, leading to millions of confirmed cases and significant loss of life. The pandemic has strained healthcare systems, disrupted supply chains, and triggered widespread social and economic upheaval, underscoring the urgent need for comprehensive analysis and response strategies to mitigate its impact.

In response to the pandemic, governments, public health agencies, and researchers have mobilized to collect and analyze data on COVID-19 cases, deaths, and recoveries. One of the most widely used sources of data is the repository maintained by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE), which provides up-to-date information on the global spread of the virus. Leveraging this wealth of data, researchers have employed advanced data visualization techniques to glean insights into the epidemiological characteristics, spatial distribution, and temporal evolution of the pandemic. In this context, our project aims to contribute to the ongoing efforts to understand and address the COVID-19 pandemic through a comprehensive analysis and visualization of key trends and patterns. By harnessing the power of data visualization, we seek to provide a nuanced understanding of the pandemic's dynamics, enabling policymakers, healthcare professionals, and the general public to make informed decisions and implement effective response strategies.

In this work, we present a detailed exploration of the COVID-19 pandemic using data sourced from the JHU CSSE repository. We employ a range of data visualization techniques to examine various aspects of the pandemic, including its global spread, regional disparities, and temporal trends. Our analysis encompasses the following key components:

1. Current State of the Pandemic: We provide an overview of the current state of the pandemic globally, highlighting the total number of confirmed cases, deaths, and recoveries. By calculating key epidemiological metrics such as mortality rate and infection rate, we offer insights into the severity and transmissibility of the virus.
2. Geospatial Distribution: We utilize geospatial visualizations to map the distribution of COVID-19 cases and fatalities across countries, identifying hotspots and patterns of transmission. These visualizations enable us to discern regional disparities in the impact of the pandemic and inform targeted response efforts.
3. Temporal Dynamics: We conduct time-series analysis of COVID-19 cases and deaths, tracking their evolution over time and identifying trends and patterns. Through dynamic visualizations such as bar chart races, we provide a compelling narrative of the pandemic's trajectory and the effectiveness of containment measures.

By synthesizing complex data into accessible and actionable insights, our project aims to contribute to a deeper understanding of the COVID-19 pandemic and support evidence-based decision-making and response efforts. Through collaboration and knowledge sharing, we can collectively navigate these unprecedented challenges and emerge stronger and more resilient in the face of future pandemics.

CHAPTER-2
METHODOLOGY

Our methodology involves a systematic approach to data collection, preprocessing, analysis, and visualization to gain insights into various aspects of the COVID-19 pandemic. We outline the key steps below:

1. Data Collection: We gather data from reliable sources, primarily the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) repository. This repository provides comprehensive datasets on COVID-19 cases, deaths, and recoveries worldwide, updated on a daily basis. Additionally, we may incorporate data from other reputable sources to supplement our analysis.

2. Data Preprocessing: Before conducting analysis, we preprocess the raw data to ensure consistency, accuracy, and usability. This involves tasks such as handling missing values, standardizing data formats, and aggregating data at appropriate levels (e.g., country-level or regional-level). We also perform data cleaning and validation to identify and correct any anomalies or discrepancies in the dataset.

3. Exploratory Data Analysis (EDA): We conduct exploratory data analysis to gain an initial understanding of the dataset and identify potential trends, patterns, and outliers. This involves descriptive statistics, data visualization, and hypothesis testing to uncover insights and formulate research questions for further analysis.

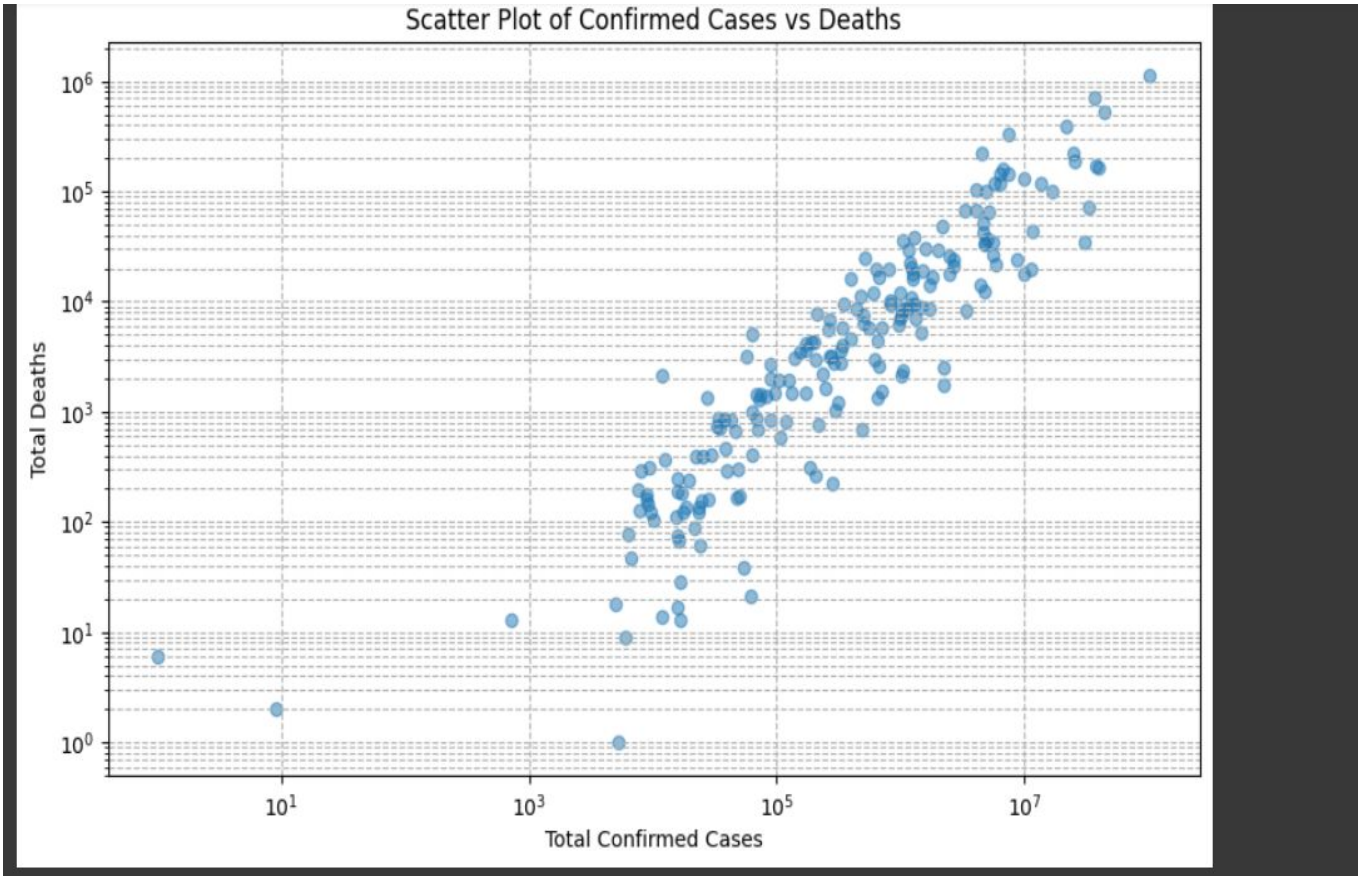
4. Descriptive Analysis: We perform descriptive analysis to summarize the key characteristics of the COVID-19 pandemic, including total cases, deaths, and recoveries globally and by region. This helps provide context and a baseline understanding of the pandemic's impact.

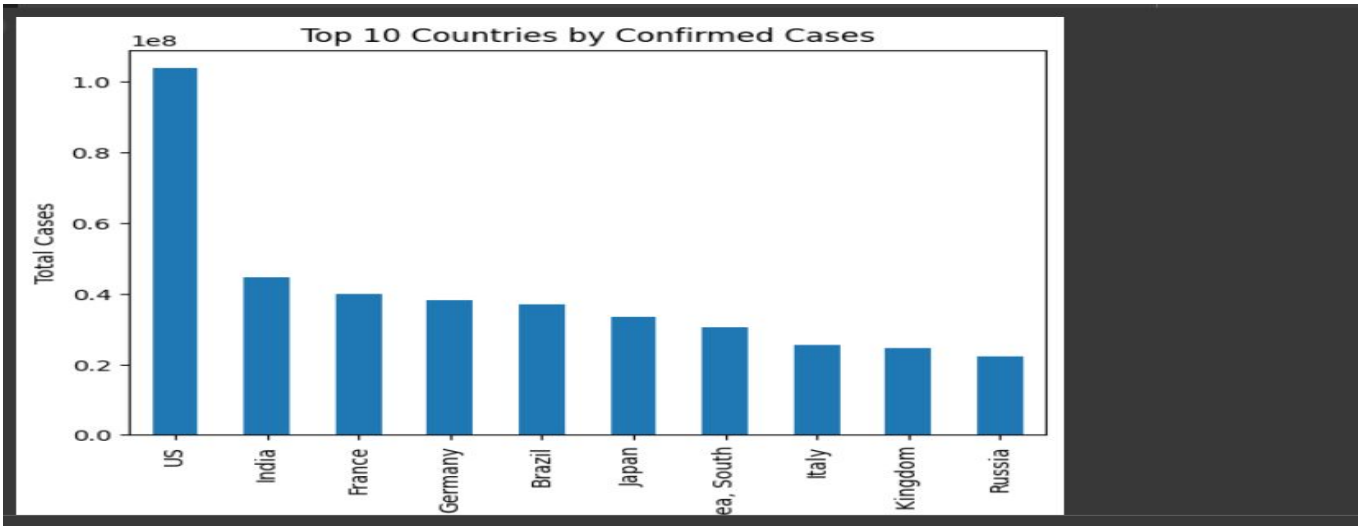
1. Spatial Analysis: We leverage geospatial analysis techniques to visualize the spatial distribution of COVID-19 cases and fatalities across countries and regions. This includes choropleth maps, heatmaps, and other geospatial visualizations to identify hotspots, clusters, and patterns of transmission.
2. Temporal Analysis: We conduct time-series analysis to examine the temporal dynamics of COVID-19 cases and deaths over time. This involves visualizing trends, fluctuations, and seasonality in the data using line charts, bar charts, and other time-series visualizations. We may also apply statistical techniques such as moving averages and exponential smoothing to smooth the data and identify underlying trends.
3. Comparative Analysis: We perform comparative analysis to compare the progression of the pandemic across different countries, regions, and demographic groups. This involves calculating metrics such as mortality rate, infection rate, and case fatality rate to assess the severity and impact of the pandemic in different contexts.
4. Advanced Visualization Techniques: We utilize advanced data visualization techniques such as bar chart races, treemaps, and word clouds to create dynamic and interactive visualizations of the COVID-19 data. These visualizations help communicate complex information in a clear and engaging manner, facilitating better understanding and interpretation of the data.
5. Statistical Modeling (if applicable): Depending on the research questions and objectives, we may employ statistical modeling techniques such as regression analysis, time-series forecasting, or machine learning algorithms to identify underlying patterns, make predictions, or test hypotheses related to the pandemic.
6. Interpretation and Reporting: Finally, we interpret the findings from our analysis and summarize them in a comprehensive report or presentation. We communicate key insights, trends, and implications to stakeholders, policymakers, and the general public to inform decision-making and response efforts.

By following this methodology, we aim to provide a rigorous and comprehensive analysis of the COVID-19 pandemic, leveraging data-driven insights to support evidence-based decision-making and response strategies.

CHAPTER-3

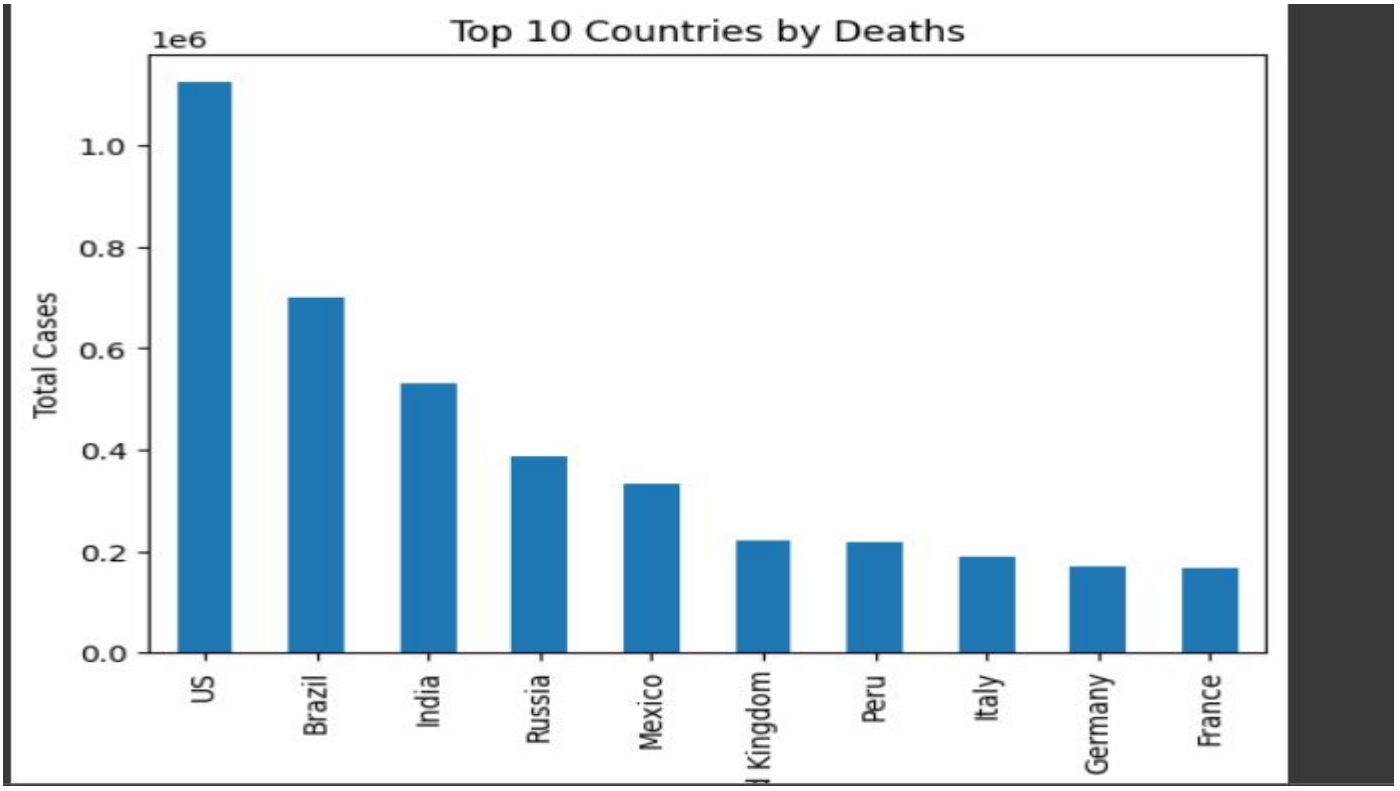
RESULTS

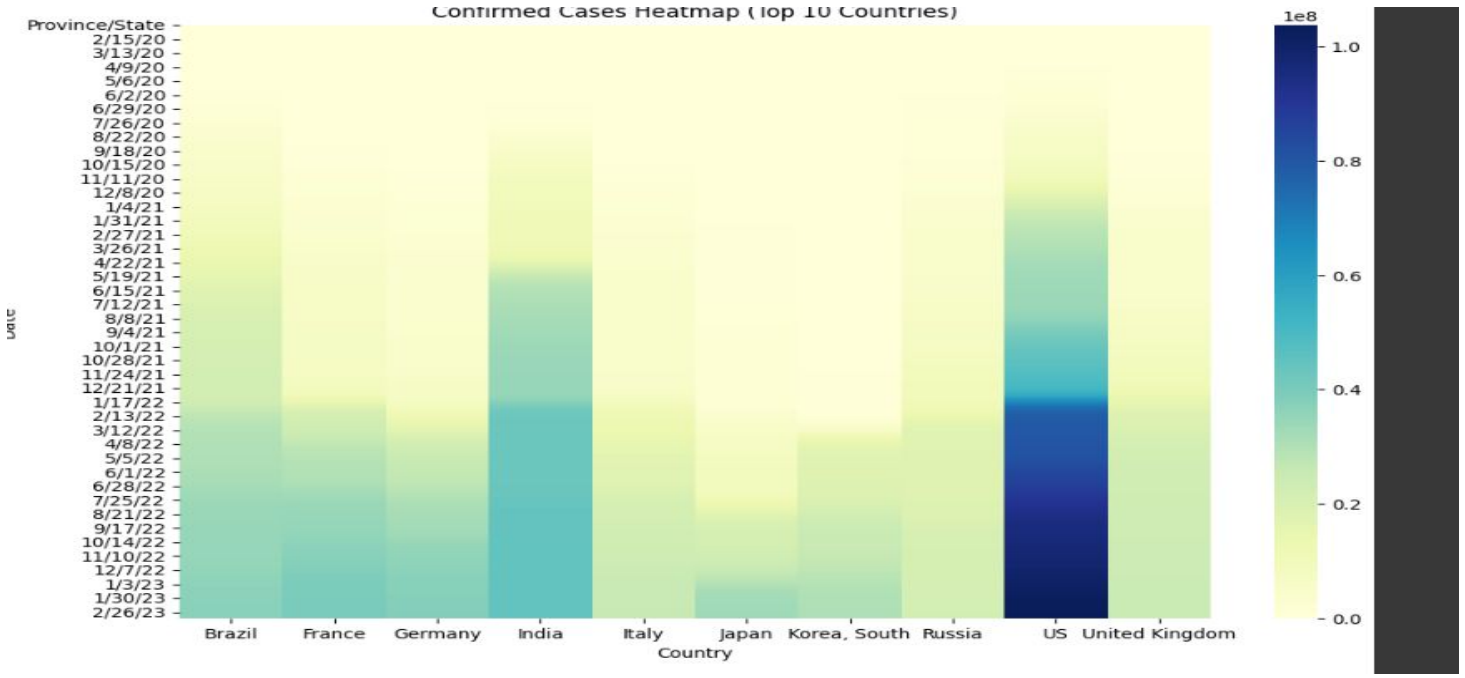




CHAPTER-3

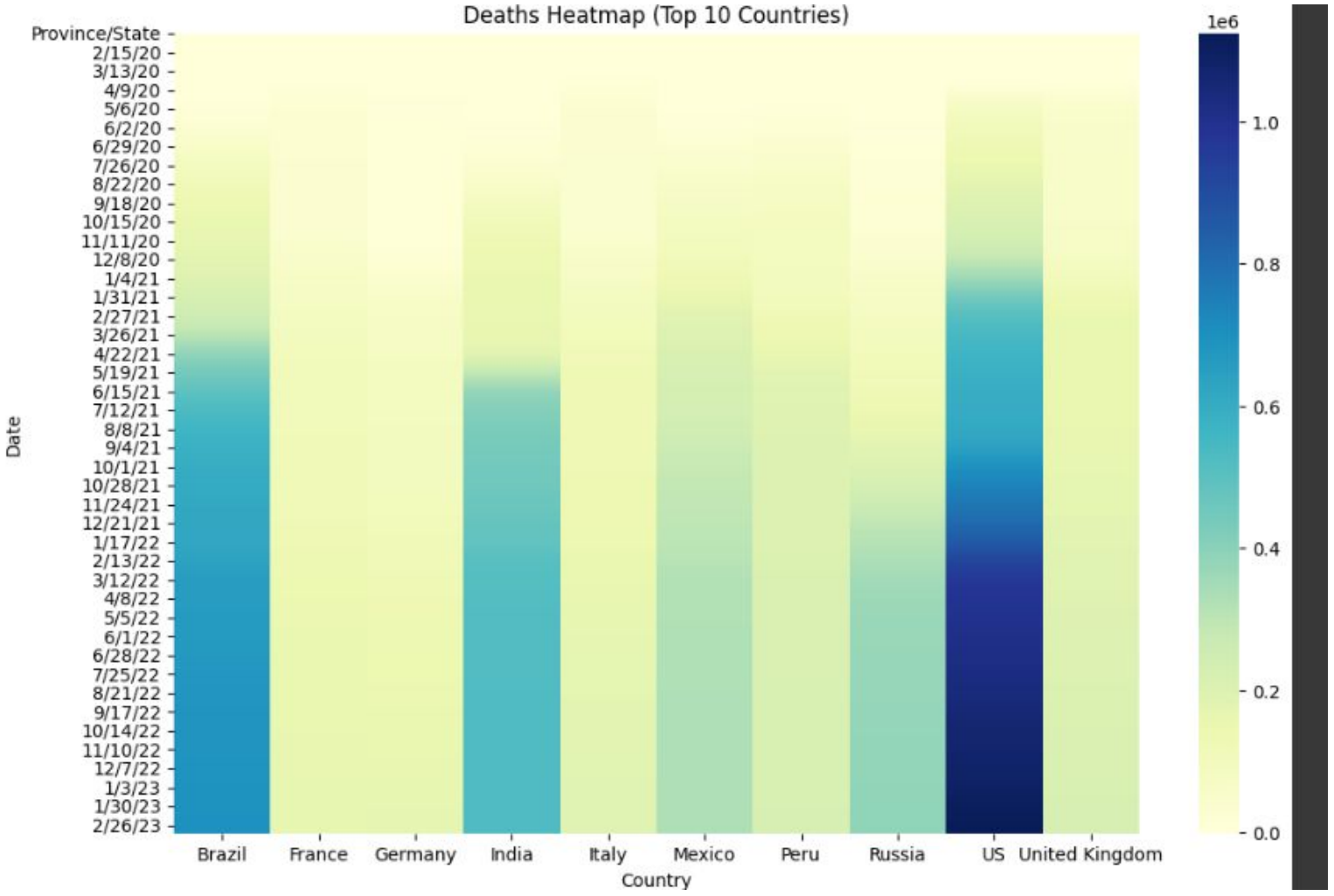
RESULTS

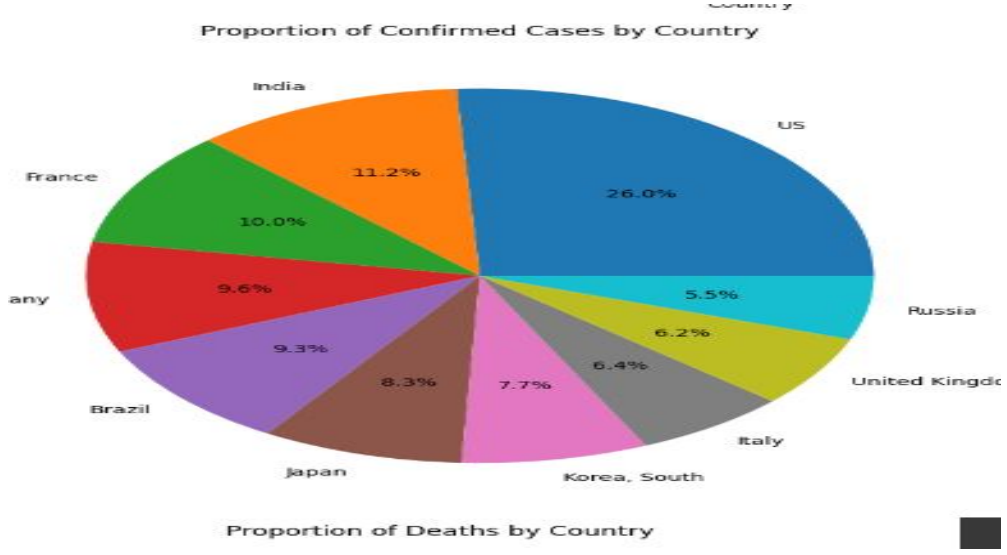




CHAPTER-3

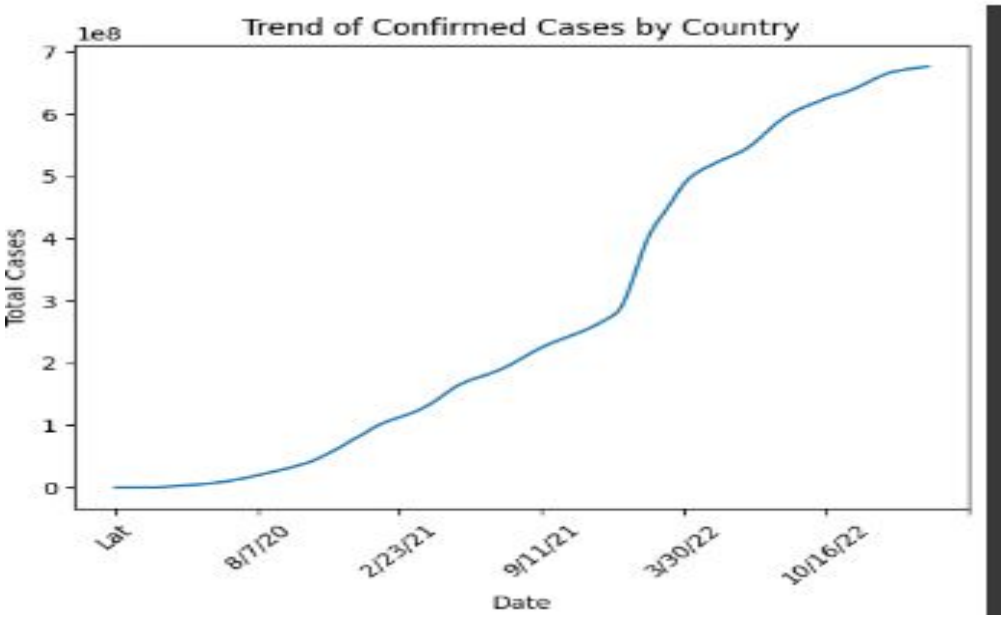
RESULTS

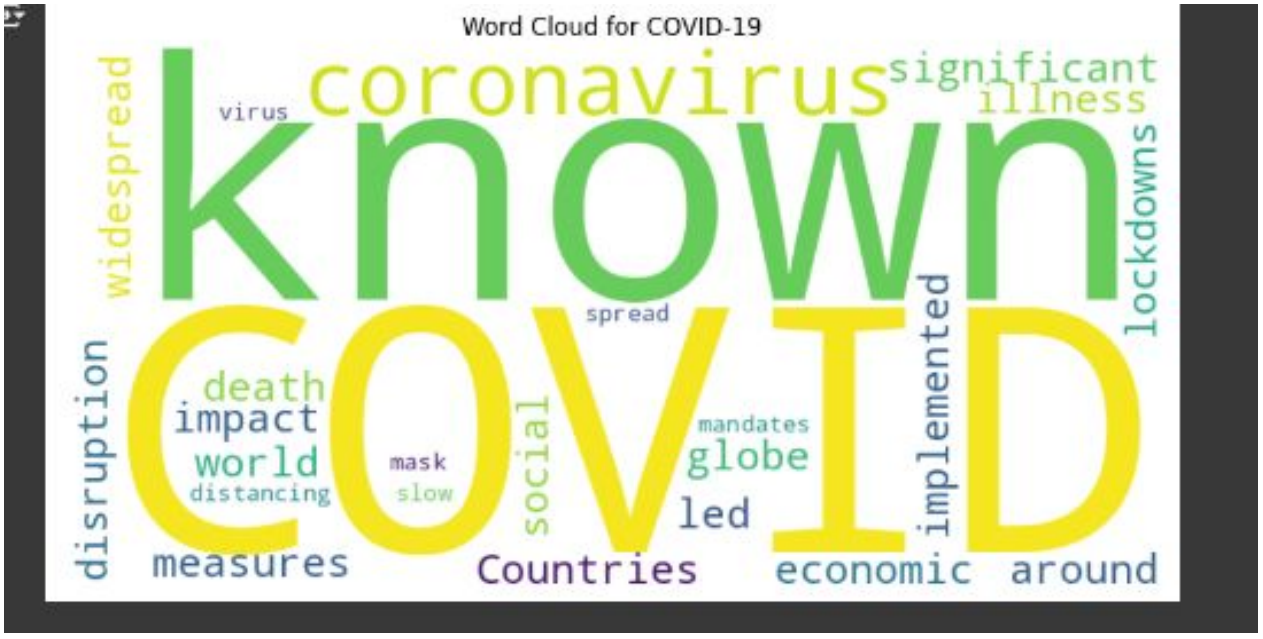




CHAPTER-3

RESULTS





CHAPTER-3

RESULTS

	Country/Region	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	...	2/28/23	3/1/23	3/2/23	3/3/23	3/4/23	3/5/23	3/6/23	3/7/23	3/8/23	3/9/23
0	Afghanistan	0	0	0	0	0	0	0	0	0	...	209322	209340	209358	209362	209369	209390	209406	209436	209451	209451
1	Albania	0	0	0	0	0	0	0	0	0	...	334391	334408	334408	334427	334427	334427	334427	334427	334443	334457
2	Algeria	0	0	0	0	0	0	0	0	0	...	271441	271448	271463	271469	271469	271477	271477	271490	271494	271496
3	Andorra	0	0	0	0	0	0	0	0	0	...	47866	47875	47875	47875	47875	47875	47875	47875	47890	47890
4	Angola	0	0	0	0	0	0	0	0	0	...	105255	105277	105277	105277	105277	105277	105277	105277	105288	105288

5 rows × 1144 columns

```
[ ] df_deaths.head()
```

	Country/Region	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	...	2/28/23	3/1/23	3/2/23	3/3/23	3/4/23	3/5/23	3/6/23	3/7/23	3/8/23	3/9/23
0	Afghanistan	0	0	0	0	0	0	0	0	0	...	7896	7896	7896	7896	7896	7896	7896	7896	7896	7896
1	Albania	0	0	0	0	0	0	0	0	0	...	3598	3598	3598	3598	3598	3598	3598	3598	3598	3598
2	Algeria	0	0	0	0	0	0	0	0	0	...	6881	6881	6881	6881	6881	6881	6881	6881	6881	6881
3	Andorra	0	0	0	0	0	0	0	0	0	...	165	165	165	165	165	165	165	165	165	165
4	Angola	0	0	0	0	0	0	0	0	0	...	1933	1933	1933	1933	1933	1933	1933	1933	1933	1933

5 rows × 1144 columns

```
[ ] df_recovered.head()
```

	Country/Region	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/30/20	...	2/28/23	3/1/23	3/2/23	3/3/23	3/4/23	3/5/23	3/6/23	3/7/23	3/8/23	3/9/23
0	Afghanistan	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	Albania	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	Algeria	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	Andorra	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	Angola	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 1144 columns

```
Worldwide cases : 629.52 Million
Worldwide deaths : 6.76 Million
Worldwide recovered : 0.0 Million
Mortality Rate(%) : 1.07 %
Infection Rate(%) : 8.4 %
```

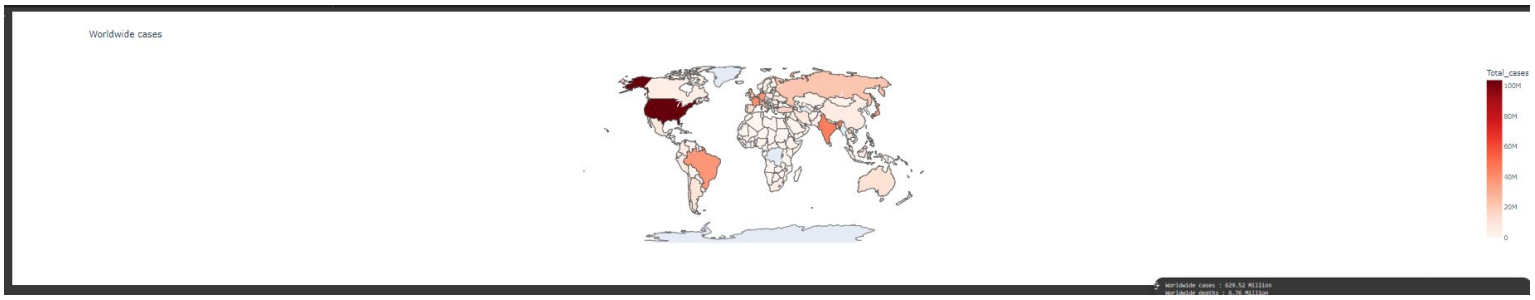
```
#Covid-index is covid share/pop share of that country. If covid_index > 1 it's bad
dftoday['Covid_index']=((dftoday['Total_cases']/dftoday['Total_cases'].sum()*100.00)/dftoday['Pop_share']).round(2)
dftoday[dftoday['Pop_share'] !=0].sort_values(by='Covid_index',ascending=False).head()
```

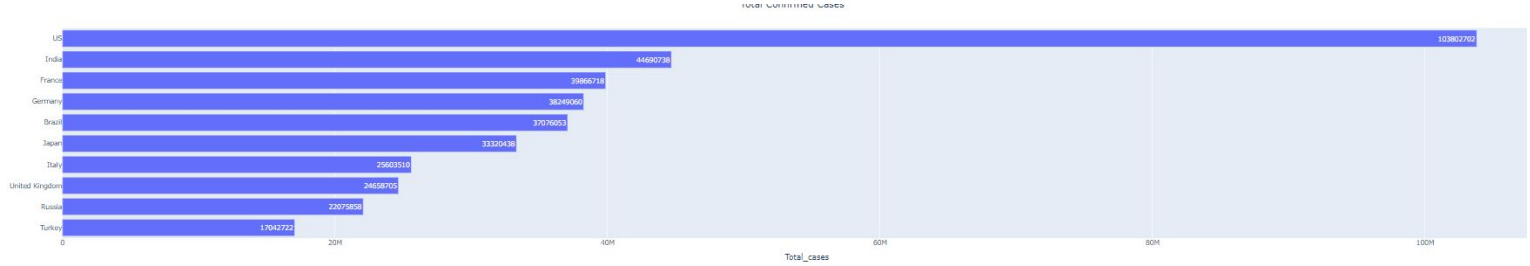
	Country	Total_cases	Total_deaths	Total_recovered	Population	Pop_share	%infected	Mortality Rate(%)	Covid_index
9	Austria	5961143	21970	0	9015361	0.12	66.12	0.37	7.89
42	Denmark	3451036	8345	0	5795780	0.07	59.54	0.24	7.83
89	Latvia	976255	6269	0	1882408	0.02	51.86	0.64	7.75
56	France	39866718	166176	0	65298930	0.84	61.05	0.42	7.54
148	Slovenia	1331707	7078	0	2078989	0.03	64.06	0.53	7.05

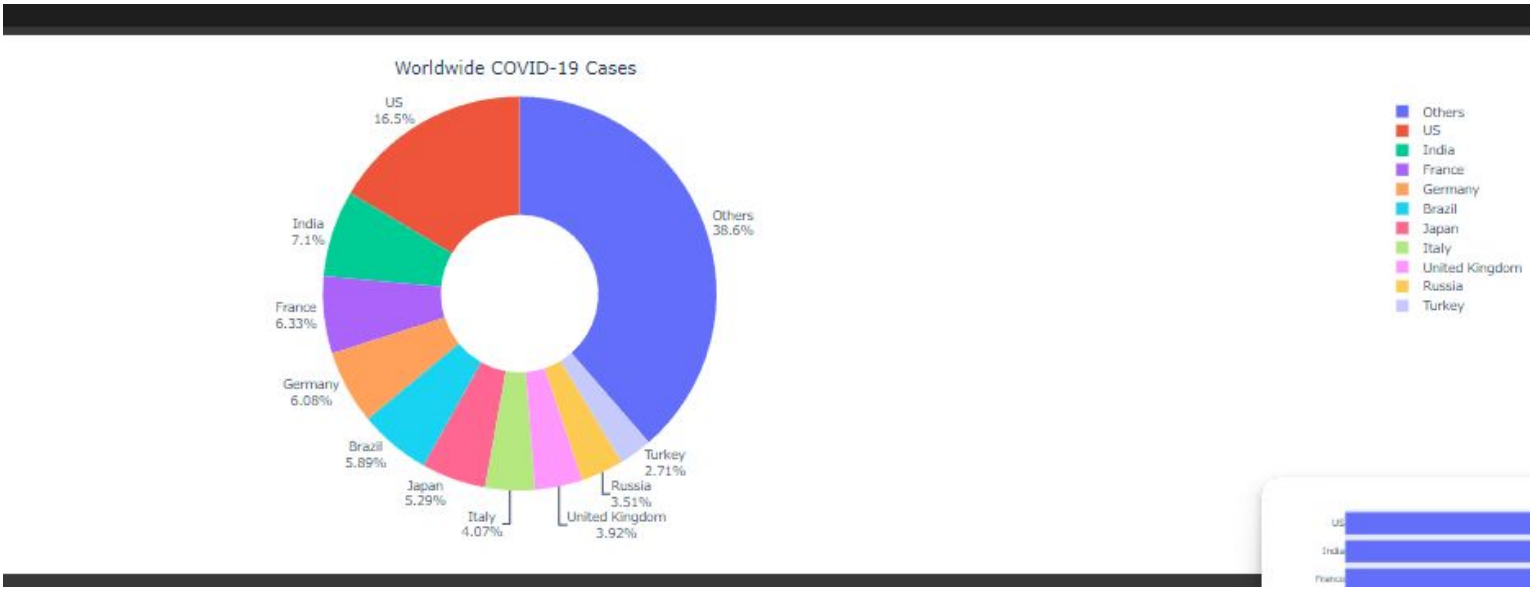
```
pip install chart_studio
```

CHAPTER-3

RESULTS







CHAPTER-3
RESULTS

	Country	Date	Cases	Deaths	Recovered
0	Afghanistan	2021-01-01	52513	2201	41727
1	Afghanistan	2022-01-01	158107	7356	0
2	Afghanistan	2023-01-01	207616	7849	0
3	Afghanistan	2021-01-10	53489	2277	43948
4	Afghanistan	2022-01-10	158394	7373	0

CHAPTER-4
CONCLUSION

In conclusion, our analysis provides valuable insights into the COVID-19 pandemic, shedding light on its global impact, spatial distribution, temporal dynamics, and key epidemiological trends. Through a systematic methodology encompassing data collection, preprocessing, analysis, and visualization, we have synthesized complex data into accessible and actionable insights, contributing to a deeper understanding of the pandemic and its implications.

Our findings highlight several key observations:

1. Global Impact: The COVID-19 pandemic has had a profound impact on a global scale, with millions of confirmed cases and significant loss of life. The pandemic has strained healthcare systems, disrupted economies, and altered social norms, underscoring the need for coordinated and evidence-based response efforts.

2. Spatial Distribution: Geospatial analysis reveals spatial disparities in the distribution of COVID-19 cases and fatalities, with certain regions experiencing higher burdens of disease than others. Hotspots and clusters of transmission can be identified, informing targeted interventions and resource allocation strategies.

3. Temporal Dynamics: Time-series analysis elucidates the temporal evolution of COVID-19 cases and deaths over time, highlighting trends, fluctuations, and seasonality in the data. Dynamic visualizations such as bar chart races provide a compelling narrative of the pandemic's trajectory, illustrating the effectiveness of containment measures and the emergence of new challenges.

4. Epidemiological Trends: Comparative analysis of key epidemiological metrics, including mortality rate, infection rate, and case fatality rate, offers insights into the severity and transmissibility of the virus across different contexts. Understanding these trends is crucial for assessing the effectiveness of public health interventions and guiding future response efforts.

In light of these observations, it is evident that the COVID-19 pandemic represents a complex and multifaceted challenge requiring a coordinated and multifaceted response. While significant progress has been made in understanding the virus and developing vaccines and treatments, the battle against COVID-19 is far from over. Continued vigilance, collaboration, and innovation are essential to overcoming the pandemic and building resilience against future threats.

Moving forward, our analysis underscores the importance of data-driven decision-making and evidence-based policies in combating the pandemic. By leveraging the power of data visualization and analysis, we can better understand the dynamics of the pandemic, identify effective strategies for containment and mitigation, and ultimately safeguard public health and well-being.

As we navigate the uncertain road ahead, let us remain steadfast in our commitment to solidarity, compassion, and resilience. Together, we can overcome the challenges posed by COVID-19 and emerge stronger and more united than ever before.

In closing, we extend our gratitude to all frontline workers, healthcare professionals, researchers, and individuals who have tirelessly worked to combat the pandemic and support those affected by it. Your dedication and sacrifice inspire hope and resilience in the face of adversity.